

日本国特許庁
JAPAN PATENT OFFICE

JC872 U.S. PRO
09/892465
06/28/01


別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application:

2000年 6月30日

出願番号
Application Number:

特願2000-200241

出願人
Applicant(s):

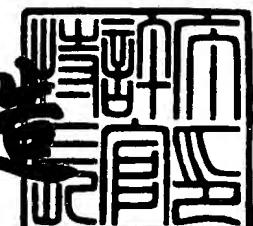
株式会社東芝

CERTIFIED COPY OF
PRIORITY DOCUMENT

2001年 4月20日

特許庁長官
Commissioner,
Japan Patent Office

及川耕造



【書類名】 特許願
【整理番号】 A000003312
【提出日】 平成12年 6月30日
【あて先】 特許庁長官 殿
【国際特許分類】 G06K 9/00
【発明の名称】 文書認識装置及びその方法
【請求項の数】 13
【発明者】
【住所又は居所】 東京都青梅市末広町2丁目9番地 株式会社東芝青梅工
場内
【氏名】 富樫 雄一
【発明者】
【住所又は居所】 東京都青梅市末広町2丁目9番地 株式会社東芝青梅工
場内
【氏名】 土内 崇靖
【特許出願人】
【識別番号】 000003078
【氏名又は名称】 株式会社 東芝
【代理人】
【識別番号】 100058479
【弁理士】
【氏名又は名称】 鈴江 武彦
【電話番号】 03-3502-3181
【選任した代理人】
【識別番号】 100084618
【弁理士】
【氏名又は名称】 村松 貞男
【選任した代理人】
【識別番号】 100068814

【弁理士】

【氏名又は名称】 塚井 淳

【選任した代理人】

【識別番号】 100092196

【弁理士】

【氏名又は名称】 橋本 良郎

【選任した代理人】

【識別番号】 100091351

【弁理士】

【氏名又は名称】 河野 哲

【選任した代理人】

【識別番号】 100088683

【弁理士】

【氏名又は名称】 中村 誠

【選任した代理人】

【識別番号】 100070437

【弁理士】

【氏名又は名称】 河井 将次

【手数料の表示】

【予納台帳番号】 011567

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 文書認識装置及びその方法

【特許請求の範囲】

【請求項1】 認識の対象となる被文書体の一部を連続的に撮像する手段と

前記撮像された複数の被文書体の画像のうち特定の被文書体の画像の文字列の画像から、比較の対象となる被文書体の画像の文字列の画像のずれ量を撮像された被文書体の画像ごとに算出する手段と、

前記算出されたずれ量が所定の量に達した場合に、前記特定の被文書体の画像の文字列の画像に、ずれ量が所定の値に達した被文書体の画像の文字列の画像のうち前記特定の被文書体の画像の文字列の画像に対して新たな文字の画像を張り合わせることにより文書画像を生成する手段と

を具備することを特徴とする文書認識装置。

【請求項2】 前記連続的に撮像された被文書体の一部の画像を格納する手段をさらに具備することを特徴とする請求項1記載の文書認識装置。

【請求項3】 前記ずれ量を算出する手段は、

前記比較の対象となる被文書体の画像のうち文字列の画像が記載されている行領域を求める手段と、

前記求められた行領域の輝度の特徴を示す行特徴射影データを求める手段と、

前記特定の被文書体の画像の行特徴射影データと、前記比較の対象となる被文書体の画像の行特徴射影データに基づいて、前記特定の被文書体の画像の文字列の画像から、比較の対象となる被文書体の画像の文字列の画像のずれ量を算出する手段と

を具備することを特徴とする請求項1記載の文書認識装置。

【請求項4】 前記ずれ量を算出する手段は、

前記比較の対象となる被文書体の画像のうち文字列の画像が記載されている列領域を求める手段と、

前記求められた列領域の輝度の特徴を示す列特徴射影データを求める手段と、

前記特定の被文書体の画像の列特徴射影データと、前記比較の対象となる被文

書体の画像の列特徴射影データとに基づいて、前記特定の被文書体の画像の文字列の画像から、比較の対象となる被文書体の画像の文字列の画像のずれ量を算出する手段と

を具備することを特徴とする請求項1記載の文書認識装置。

【請求項5】 前記所定の値は、前記特定の被文書体の画像の行特徴射影データの形状に基づいて決定されることを特徴とする請求項1記載の文書認識装置。

【請求項6】 前記連続的に撮像された認識の対象となる複数の被文書体の一部の画像をそれぞれ表示する手段をさらに具備することを特徴とする請求項1記載の文書認識装置。

【請求項7】 前記生成された文書画像を第1の文書データに変換する手段と、

前記変換された第1の文書データを表示する手段と、

前記表示された第1の文書データに基づいて、前記撮像手段によって前記認識の対象となる被文書体の一部がズームアップされて撮像された場合に、前記ズームアップされて撮像された被文書体の一部の画像データを第2の文書データに変換する手段と、

前記第1の文書データのうち前記第2の文書データと異なる文字を、前記第2の文書データのうち前記異なる文字と対応する文字に置き換える手段とをさらに具備することを特徴とする請求項1記載の文書認識装置。

【請求項8】 認識の対象となる被文書体の一部を連続的に撮像し、

前記撮像された複数の被文書体の画像のうち特定の被文書体の画像の文字列の画像から、比較の対象となる被文書体の画像の文字列の画像のずれ量を撮像された被文書体の画像ごとに算出し、

前記算出されたずれ量が所定の量に達した場合に、前記特定の被文書体の画像の文字列の画像に、ずれ量が所定の値に達した被文書体の画像の文字列の画像のうち前記特定の被文書体の画像の文字列の画像に対して新たな文字の画像を張り合わせることにより文書画像を生成することを特徴とする文書認識方法。

【請求項9】 前記ずれ量を算出するステップは、

前記比較の対象となる被文書体の画像のうち文字列の画像が記載されている行領域を求め、

前記求められた行領域の輝度の特徴を示す行特徴射影データを求め、

前記特定の被文書体の画像の行特徴射影データと、前記比較の対象となる被文書体の画像の行特徴射影データとに基づいて、前記特定の被文書体の画像の文字列の画像から、比較の対象となる被文書体の画像の文字列の画像のずれ量を算出することを特徴とする請求項8記載の文書認識方法。

【請求項10】 前記ずれ量を算出するステップは、

前記比較の対象となる被文書体の画像のうち文字列の画像が記載されている列領域を求め、

前記求められた列領域の輝度の特徴を示す列特徴射影データを求め、

前記特定の被文書体の画像の列特徴射影データと、前記比較の対象となる被文書体の画像の列特徴射影データとに基づいて、前記特定の被文書体の画像の文字列の画像から、比較の対象となる被文書体の画像の文字列の画像のずれ量を算出することを特徴とする請求項8記載の文書認識方法。

【請求項11】 前記所定の値は、前記特定の被文書体の画像の行特徴射影データの形状に基づいて決定されることを特徴とする請求項8記載の文書認識方法。

【請求項12】 前記連続的に撮像された認識の対象となる複数の被文書体の一部の画像をそれぞれ表示する手段をさらに具備することを特徴とする請求項8記載の文書認識方法。

【請求項13】 前記生成された文書画像を第1の文書データに変換し、

前記変換された第1の文書データを表示し、

前記表示された第1の文書データに基づいて、前記撮像手段によって前記認識の対象となる被文書体の一部がズームアップされて撮像された場合に、前記ズームアップされて撮像された被文書体の一部の画像データを第2の文書データに変換し、

前記第1の文書データのうち前記第2の文書データと異なる文字を、前記第2の文書データのうち前記異なる文字と対応する文字に置き換えるステップをさら

に具備することを特徴とする請求項8記載の文書認識方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、文書認識装置及びその方法に関する。

【0002】

【従来の技術】

従来より、文字を認識する装置としては、OCR (Optical Character Reader) が広く知られている。

【0003】

このようなOCRにおいては、文書画像をCCD密着センサを用いたスキャナで読み込み文書データを求めている。CCD密着センサで読み取られた画像は、2値化処理・文字切り出し処理・文字正規化が行われた2値データとなり、文字辞書とマッチング処理を行なうことにより文字データを得る。

【0004】

さらに、文書には複数の文字が書かれているので、連続する複数の文字から単語あるいは文書の形式に照らしあわせて文書データとしていた。

【0005】

また、OCRの他に、カメラを用いて画像を撮り、その画像中の文字を認識することが考えられている。しかしながら、カメラのCCDは、動画を撮ることを目的としているために解像度はスキャナに比較して高くない。

【0006】

そのため、文書全体を撮像すると文字が小さくなり、文字の認識率に影響を及ぼすために、ズームアップして撮像する。すると1度に読める文字の数が少なく文書としての認識が難しいといった問題点があった。

【0007】

そのために、複数枚画像を撮って張り合わせる手法が考えられている。自然画像などで行われている手法で、画像中の特徴を捕らえて同じ部分をオーバーラップさせるように画像を張り合わせる。その画像を用いて文字認識させることができ

きるが、従来の技術ではつなぎめでの文字を読み誤る可能性があった。

【0008】

また、認識結果において誤った場合には、通常その文字をキーボード及びマウスなどで選択して、修正する処理を加えることが行われてきた。

【0009】

【発明が解決しようとする課題】

従来のOCRを使用して文書を認識しようとする場合、スキャナで用いられている密着型CCDで画像を取り込んでいたために、読み取り対象文書を平らな台上に置くあるいは、1枚づつ切り離して読み込ませる必要があった。そのために、壁に貼られた紙の文字を認識処理を行ないにくいという問題があった。

【0010】

また、カメラを使用して文書を認識しようとする場合、一般的なTVカメラから画像を取り込んだ時の解像度は 640×480 画素であり、画像全体を一度で読み込むと1文字のデータ量が少なすぎるために認識性能が低いという問題があった。

【0011】

また、ズームアップして1文字あたりのデータを増やすと、少ない領域の画像しか得ることができないので、一度に読む文字数が限られる。すると、単語あるいは日本語の構造情報を用いたあと処理ができないので認識率が良くならないという問題があった。

【0012】

また、複数枚の画像をつなぎ合わせると、つなぎめの文字を読みあやまることなどの問題点や、つながらないような画像を撮影してしまうなどの問題点があった。

【0013】

さらに、カメラを用いて文字読み取りさせる時には、カメラを手で操作する必要があるために、誤った文字を修正するためにマウスやキーボードを用いると操作が煩雑になるなどのという問題があった。

【0014】

本発明は、上記実情に鑑みてなされたものであり、カメラを動かすことで、広い部分の文書を高精度で読み取り、かつ誤認識部分の修正を容易に行なうことのできるカメラ画像認識装置を提供することを目的とする。

【0015】

また、本発明は、カメラを動かすことで、広い部分の文書を高精度で読み取り、かつ誤認識部分の修正を容易に行なうことのできるカメラ画像認識方法を提供することを目的とする。

【0016】

【課題を解決するための手段】

したがって、まず、上記目的を達成するために、本発明の内第1の発明の文書認識装置は、認識の対象となる被文書体の一部を連続的に撮像する手段と、前記撮像された複数の被文書体の画像のうち特定の被文書体の画像の文字列の画像から、比較の対象となる被文書体の画像の文字列の画像のずれ量を撮像された被文書体の画像ごとに算出する手段と、前記算出されたずれ量が所定の量に達した場合に、前記特定の被文書体の画像の文字列の画像に、ずれ量が所定の値に達した被文書体の画像の文字列の画像のうち前記特定の被文書体の画像の文字列の画像に対して新たな文字の画像を張り合わせることにより文書画像を生成する手段とを具備することを特徴とする。

【0017】

このような発明によれば、カメラで画像をスキャンすることで解像度の高い画像を得て文字の読み取りができる。また、行の途中まで文章を読みたいような場合に、途中までの画像を入力してインタラクティブに文書を読み取ることができる。

【0018】

さらに、本発明の内第2の発明の文書認識装置は、第1の発明において、前記連続的に撮像された認識の対象となる複数の被文書体の一部の画像をそれぞれ表示する手段をさらに具備することを特徴とする。

【0019】

このような発明によれば、カメラで複数の画像を取り込む際に、つなぎあわせ

るのに最適な画像の取り込みをすることができる。

【0020】

さらに、本発明の第3の発明の文書認識装置は、第1の発明において、前記生成された文書画像を第1の文書データに変換する手段と、前記変換された第1の文書データを表示する手段と、前記表示された第1の文書データに基づいて、前記撮像手段によって前記認識の対象となる被文書体の一部がズームアップされて撮像された場合に、前記ズームアップされて撮像された被文書体の一部の画像データを第2の文書データに変換する手段と、前記第1の文書データのうち前記第2の文書データと異なる文字を、前記第2の文書データのうち前記異なる文字と対応する文字に置き換える手段とをさらに具備することを特徴とする。

【0021】

このような発明によれば、被文書体の一部をカメラでズームアップするだけで、簡単に誤認識された文字を修正することができる。

【0022】

【発明の実施の形態】

以下、図面を参照して本発明の実施の形態について説明する。

【0023】

＜第1の実施の形態＞

図1は、本発明の第1の実施の形態に係る文書認識装置の構成を示す図である。

【0024】

同図に示すように、本実施の形態の文書認識装置は、カメラ1、A/D変換器2、画像メモリ3、D/A変換器4、表示機5及びCPU6を有している。

【0025】

カメラ1は、被写体である文書を撮像し、この撮像した文書を示す文書画像データをA/D変換器2及び表示機5に出力する。なお、カメラ1は、動画像を撮影するためのTVカメラ、静止画像を撮影するためのスチルカメラのいずれであってもよい。

【0026】

A/D変換器2は、カメラ1から出力される文書画像データをデジタル信号に変換し画像メモリ3に出力する。

【0027】

画像メモリ3は、A/D変換器2から出力される文書画像データを格納する。具体的には、画像メモリ3は、カメラ1により連続的に撮像される複数の画像を格納するものであり、後に述べるマスター文書画像や比較対象文書画像などを格納するものである。

【0028】

D/A変換器4は、画像メモリ3に格納された文書画像データをアナログ信号に変換して表示機5に出力する。

【0029】

表示機5は、D/A変換器4から出力される文書画像データ及びカメラ1から出力される文書画像を表示する。

【0030】

CPU6は、A/D変換器2、画像メモリ3及びD/A変換器4など全体の制御を司る。具体的には、図4、図6及び図17に示すフローチャートの処理などを行なう。

【0031】

本実施の形態は、文書認識装置において画像を取り込む場合に、図2に示すようにカメラ1を文書の書いてある被写体10と平行方向に移動させて連続画像を取り込み、その連続画像をつなぎ合わせた画像を生成し、そこに書かれている文字あるいは文章を読み取る方法を示したものである。

【0032】

以下、本発明の実施の形態に係る文書認識装置の動作を図4のフローチャートを参照して説明する。

【0033】

まず、最初に、ユーザは、カメラ1を手にとり、被文書体に記載された文書を撮像する。図3は、横書きの被写体文書10に対してカメラ1を左から右に動かして文書全体を撮像する様子を示す図である。

【0034】

同図においては、1枚目からn枚目のカメラ撮像範囲1～Xの画像が撮像される場合を示している。カメラ1において撮像されたカメラ撮像範囲における画像は、A/D変換器2によってデジタル信号に変換された後に、画像メモリ3に順次格納される。

【0035】

また、カメラ1における撮像と並行して、まず、1枚目の画像の取り込みが行なわれる(S1)。図3に示した場合を例にとると、被写体文書10のうち最も左側にあるカメラ撮像範囲Aの画像が取り込まれる。

【0036】

この1枚目の画像は、マスターの文書画像となり、このマスターの文書画像を基準に、後に述べるずれ量の算出や画像合成処理が行なわれる。なお、本実施の形態においては、1枚目の文書画像をマスター文書として説明するが、マスター文書とは基準となる画像という意味なので、1枚目の画像に限らず、任意の画像をマスター文書とすることができる。

【0037】

次に、取り込まれた1枚目の文書(マスター文書)に対して行領域の検出が行なわれる(S2)。

【0038】

この行領域の検出について、図5及び図6のフローチャートを参照して説明する。

【0039】

まず、取り込まれた1枚目の文書(マスター文書)に対して垂直射影データV(y)の算出が行なわれる(S11)。

【0040】

この垂直射影データV(y)の算出は、図5に示すように、行方向に輝度データを足しこむ(V軸に向かって足しこむ)ことにより得られる。図5に示すように、行の場所においては、文字データが多いので山となる。逆に行間は文字が少なく谷になる。

【0041】

ここで、垂直射影データV(y)は、式(1)で示される。

【0042】

【数1】

$$V(y) = \sum_{x=0}^n P_i x(x, y) \cdots \cdots \text{ (式1)}$$

【0043】

なお、 $P_i x(x, y)$ は、x, y座標の輝度値を表わす。

【0044】

次に、算出された垂直射影データV(y)のうち比較の対象となる垂直射影データV(y)、例えば、垂直射影データV(0)が所定のしきい値よりも大きいか否かの判断が行なわれる(S12)。

【0045】

S12において、比較の対象となる垂直射影データV(y)が所定のしきい値よりも大きいと判断された場合には、当該部分は行領域であると判断される。一方、S12において、比較の対象となる垂直射影データV(y)が所定のしきい値よりも小さいと判断された場合には、当該部分は行領域ではないと判断される(S15)。

【0046】

そして、行領域の検出が終了したか否かの判断が行なわれる(S14)。具体的には、算出された垂直射影データV(y)のy方向の全てについて行領域であるか否かの判断が行なわれた場合に行領域の検出処理が終了する。図5においては、YS0~YE0、YS1~YE1の部分が行である。

【0047】

次に、S3において、求められた行領域の行特徴射影データを算出する。この行特徴射影データは、2枚目以降の文書画像データとのマッチングのために使用される。また、算出された行特徴射影データに基づいて、文字のない間隔を求める。

【0048】

ここで、「文字のない間隔」とは、文字間隔と似ている概念であるが、文字間隔は文字と文字との間の間隔を示すものであるのに対し、文字のない間隔とは、文字のない部分（空白部分）と文字のない部分との間隔を示すものである点において異なる。

【0049】

行特徴射影データは、図7に示すように、1行の画像を行方向と垂直に画素データを足しこんだデータのことである。A/D変換器により、黒い画素が255から白い画素が0になるような連続値でA/D変換するとすると、黒い部分、すなわち文字が書かれている部分のデータが足しこまれると行特徴射影データは山になり、白い部分、すなわち文字のない部分の行特徴射影データは谷になる。このデータを検出された行ごとに求める。また、この求められた行特徴射影データに基づいて、文字のない間隔を算出する。

【0050】

具体的には、行特徴射影データは、下記式(2)にて表わされる。

【0051】

【数2】

$$\text{Proj}(n, x) = \sum_{y=y_s}^{y_e} \text{Pix}(x, y) \quad \dots \dots \dots \quad (\text{式2})$$

【0052】

次に、次の画像（2枚目の文書画像）を取り込む（S4）。そして、求められた次の文書画像の行領域の検出が行なわれる（S2）。この行領域の検出処理は、S2において述べた処理と同様である。

【0053】

次に、検出された行領域から行特徴射影データの算出が行なわれる（S6）。この行特徴射影データの算出処理は、S3において述べた処理と同様である。

【0054】

次に、1枚目の文書画像（マスター文書画像）と、取り込まれた文書画像（比較対象文書画像）とがどれだけずれているかを示すずれ量を算出する。

【0055】

なお、ここでは、マスター文書は1枚目の文書画像であるとして説明しているが、1枚目の文書画像に限らず、基準となる文書画像であればよい。

【0056】

このずれ量の算出は、マスター文書画像から得られる行特徴射影データと、比較対象文書画像から得られる行特徴射影データとのずれ量から求められる。

【0057】

具体的には、比較対象文書画像から得られた行特徴射影データをずらしながらマスター文書画像から得られた行特徴射影データに対してマッチング処理を行なう。

【0058】

仮に、カメラが+X画素分だけ移動して画像を撮っていれば-Xずらした場合にマッチングされる。なお、上記説明では、比較対象文書画像をずらしてマッチング処理を行なう場合について説明したが、マスター文書画像から得られる行特徴射影データをずらして比較対象文書画像の行特徴射影データにマッチング処理を行なってもよいことはいうまでもない。

【0059】

マッチング処理は、マスター文書画像の行特徴射影データの各度数（行特徴射影データの値）と、比較対象文書画像の行特徴射影データの各度数（行特徴射影データの値）との差を足して求める。そして、この値が最も小さい場合、マッチングされているとみなすことができる。

【0060】

具体的には、マッチング処理における差分の算出は、下記式(3)にて表わされる。

【0061】

【数3】

$$D1st = \min \left(\sum_p proj(n, x-p) - proj(n+1, x) \right) \dots \dots \dots \text{ (式3)}$$

【0062】

マッチング処理において、マスター文書画像の行特徴射影データと比較対象文

書画像の行特徴射影データとがマッチングした場合、マッチングした時の比較対象文書画像（或いはマスター文書画像）のずらし量からずれ量が検出される（S7）。

【0063】

そして、次に、この検出されたずれ量が文字のない間隔よりも大きいか否かの判断が行なわれる（S8）。S8において、ずれ量が文字のない間隔よりも少ないと判断された場合には、再度、S4の処理に移り、次の画像に対してずれ量の検出が行なわれる。なお、文字のない間隔は、図7に示すように、行特徴射影データの谷の間隔で求められる。

【0064】

一方、S8において、ずれ量が文字のない間隔よりも大きいと判断された場合には、画像合成処理（S9）に移る。以下、画像合成処理について説明する。

【0065】

図8につなぎ合わせの様子を示す。その際には、マスター文書画像に新しい部分を重ねて描画を行うが、オーバーラップした部分は、画像の焦点があつてあるかの計算をして、鮮明な方の画像を用いることにする。

【0066】

図8においては、マスター文書画像に対して「け」の文字が画像合成されることになる。

【0067】

また、V軸方向へ画像がずれて入力されることがある。このずれは、前述のV軸への射影データをとったさいに、射影の波形のずれを求めることで検出が可能である。簡単にはXE0, XE1, YB0, YB1などの値の差で求まる。より厳密に求めるには、2つのV軸射影のマッチングを行つて求めることができる。

【0068】

マッチング手法は前述の行特徴射影データのマッチングと同様の処理を行う。このずれが所定の値以内であれば、無視して張り合わせることが可能であるが、ある値以上であれば、補正して張り合わせる。また、補正できないほど大きくずれている場合には画像の張り合わせができない由の警告を発する。

【0069】

なお、本実施の形態においては、カメラ1を左から右へ動かした場合、あるいは被写体文書10を右から左へ動かした場合について説明したが、同様にカメラ1を右から左に動かす場合や、被写体文書10を左から右へ動かす場合に適応させることも可能である。

【0070】

また、本実施の形態においては、横書きの文の例について説明したが、縦書きの場合においては、カメラを上から下または下から上へ移動させて撮影すれば、同様の処理で画像を張り合わせることが可能である。

【0071】

なお、横書き縦書きの認識は、図9及び図10に示すように、画面全体をV軸、H軸に射影データを求め、波の大きさで判断することができる。

【0072】

したがって、本実施の形態においては、ある特定範囲だけの文字を読みたい場合などは、その領域のみカメラを移動させて、張り合わせの結果画像を見ながらインタラクティブに認識させることができるものとなる。

【0073】

＜第2の実施の形態＞

次に、本発明の第2の実施の形態について説明する。

【0074】

本実施の形態の文書認識装置は、第1の実施の形態の文書認識装置において、カメラによって撮像された画像をつなぎ合わせて表示するものである。

【0075】

図11は、4枚の画像をつなぎ合わせて表示する例を示している。この画像の繋ぎ合わせは、具体的には、すでに撮像された画像を画像メモリ3から読み出し、D/A変換器4を通して表示機5に表示するとともに、新たに撮像された画像を同時に表示を行ない、撮像画像の基準とする。

【0076】

したがって、本実施の形態の文書認識装置によれば、すでに撮像された画像を

表示することができるので、ユーザがカメラを移動する場合に、表示された画像を参照しながら撮影を行なうことができる。

【0077】

＜第3の実施の形態＞

本実施の形態の文書認識装置は、第1の実施の形態の文書認識装置に加えて、文字列の中の1部の文字を誤認識した際に、当該文字をズームアップして撮影することで、解像度の高い画像を取り直し再度認識を行い、自動的に文字の修正を行なう方法である。

【0078】

以下、本実施の形態の文書認識装置の動作について、図17のフローチャートを参照して説明する。

【0079】

まず、画像合成処理によって得られた文書画像の取り込みが行なわれる（S21）。次に、取り込んだ画像に対して文字認識が行なわれ（S22）、文書を作成する。

【0080】

また、この際、レイアウト情報を併せて出力しておく。このレイアウト情報は、たとえばN行M列の文字は‘あ’のような形で出力するか、或いは右からXn m上からYnmの場所にある文字は‘あ’である形で出力してもよい。

【0081】

次に、認識された文字の表示を行なう（S23）。ここで、文書全体の画像が図12に示すような場合であると仮定する。そして、図13に示すような認識結果が得られ、表示されたものと仮定する。この場合、「サンプル」の文字が誤って「サソブル」と認識されている。

【0082】

次に、ユーザは、表示された認識結果を見て、誤認識された文字を認識し、誤認識した場所へカメラを近づけたりレンズを操作してズームアップを行ない、画像を取り込む（S24）。

【0083】

図14は、ズームアップする撮像領域を示す図であり、図15は、ズームアップされて取り込まれた画像を示す図である。

【0084】

次に、取り込まれた画像に対して文字認識を行ない（S25）、最初に認識された文字列とマッチング処理を行なう（S26）。マッチングの結果とレイアウト情報より、初回に認識した文字のどこが2回目の文字領域かを求める。もちろん、誤認識の文字情報は異なるので、完全にマッチングされることはないが、その他の文字の位置関係は一致するはずである。

【0085】

次に、最初に認識された文字列と、ズームアップして撮像された画像から認識された文字列との差を検出し（S27）、誤認識された文字を入れ替える（S28）。図16は、誤認識された「サソブル」の文字を「サンブル」の文字に置き換える場合を説明するための図である。

【0086】

したがって、本実施の形態の画像認識装置によれば、カメラにより再度誤認識された文書画像をズームアップすることにより、簡単に誤認識された文字を修正することができる。

【0087】

なお、本願発明は、上記各実施形態に限定されるものでなく、実施段階ではその要旨を逸脱しない範囲で種々に変形することが可能である。また、各実施形態は可能な限り適宜組み合わせて実施してもよく、その場合組み合わされた効果が得られる。

【0088】

さらに、上記各実施形態には種々の段階の発明が含まれており、開示される複数の構成要件における適宜な組み合わせにより種々の発明が抽出され得る。例えば実施形態に示される全構成要件から幾つかの構成要件が省略されることで発明が抽出された場合には、その抽出された発明を実施する場合には省略部分が周知慣用技術で適宜補われるものである。

【0089】

また、実施形態に記載した手法は、計算機（コンピュータ）に実行させることができるプログラム（ソフトウェア手段）として、例えば磁気ディスク（フロッピーディスク、ハードディスク等）、光ディスク（CD-ROM、DVD、MO等）、半導体メモリ（ROM、RAM、フラッシュメモリ等）等の記録媒体に格納し、また通信媒体により伝送して頒布することもできる。なお、媒体側に格納されるプログラムには、計算機に実行させるソフトウェア手段（実行プログラムのみならずテーブルやデータ構造も含む）を計算機内に構成させる設定プログラムをも含む。本装置を実現する計算機は、記録媒体に記録されたプログラムを読み込み、また場合により設定プログラムによりソフトウェア手段を構築し、このソフトウェア手段によって動作が制御されることにより上述した処理を実行する。なお、本明細書でいう記録媒体は、頒布用に限らず、計算機内部あるいはネットワークを介して接続される機器に設けられた磁気ディスクや半導体メモリ等の記憶媒体を含むものである。

【0090】

【発明の効果】

以上詳記したように、本発明によれば、カメラを動かすことで、広い部分の文書を高精度で読み取り、かつ誤認識部分の修正を容易に行なうことのできるカメラ画像認識装置を提供することができる。

【0091】

また、カメラを動かすことで、広い部分の文書を高精度で読み取り、かつ誤認識部分の修正を容易に行なうことのできるカメラ画像認識方法を提供することができる。

【図面の簡単な説明】

【図1】

本発明の第1の実施の形態に係る文書認識装置のハードウェア構成を示す図。

【図2】

同第1の実施の形態における文書認識装置により画像を取り込みを説明するための図。

【図3】

横書きの被写体文書に対してカメラを左から右に動かして文書全体を撮像する様子を示す図。

【図4】

同第1の実施の形態における文書認識装置の動作を説明するためのフローチャート。

【図5】

垂直射影データを説明するための図。

【図6】

行領域の検出動作を説明するためのフローチャート。

【図7】

行特徴射影データを示す図。

【図8】

画像のつなぎ合わせを説明するための図。

【図9】

縦書き及び横書きの判別を説明するための図。

【図10】

縦書き及び横書きの判別を説明するための図。

【図11】

4枚の画像をつなぎ合わせて表示する例を示す図。

【図12】

文書全体を示す図。

【図13】

認識結果を示す図。

【図14】

ズームアップする撮像領域を示す図。

【図15】

ズームアップされて取り込まれた画像を示す図。

【図16】

誤認識された「サソブル」の文字を「サンブル」の文字に置き換える場合を説

明するための図。

【図17】

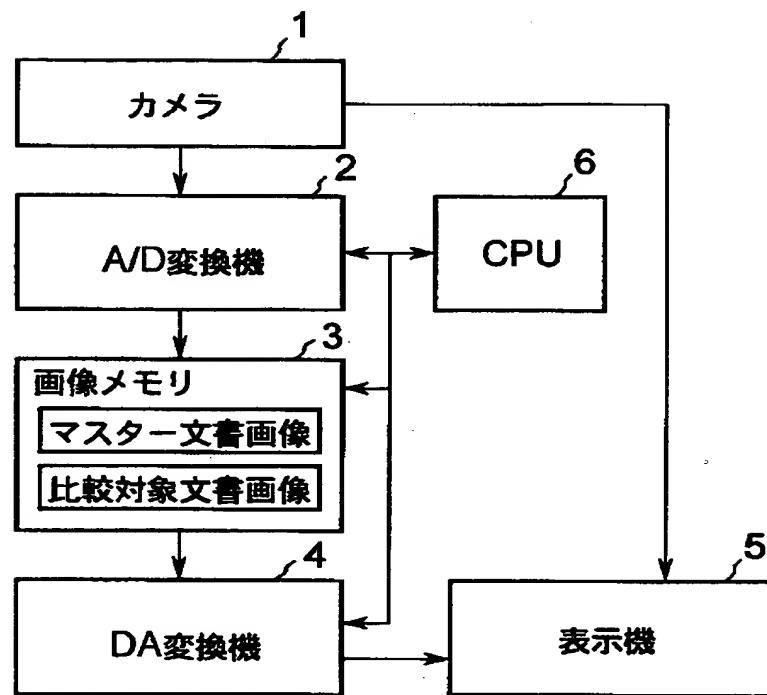
本発明の第3の実施の形態に係る文書認識装置の動作を説明するためのフローチャート。

【符号の説明】

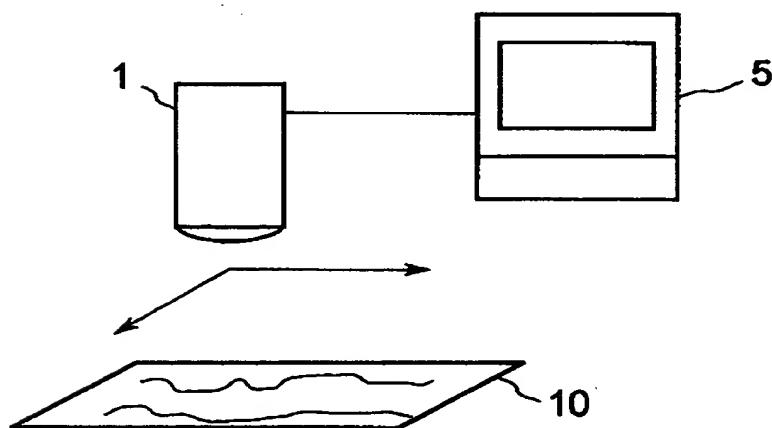
- 1 … カメラ、
- 2 … A／D変換器、
- 3 … 画像メモリ、
- 4 … D／A変換器、
- 5 … 表示機、
- 6 … C P U、
- 10 … 被写体。

【書類名】 図面

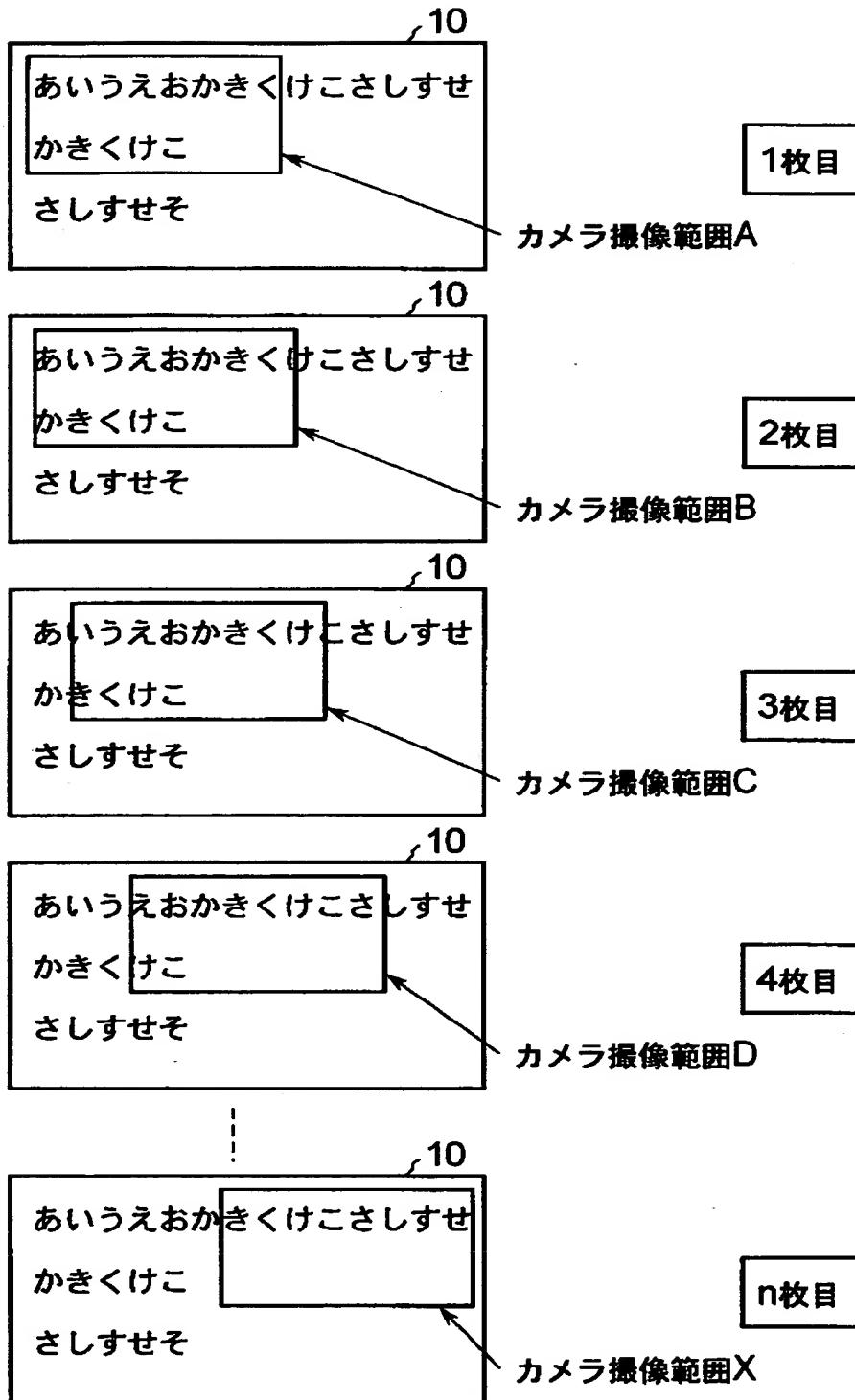
【図1】



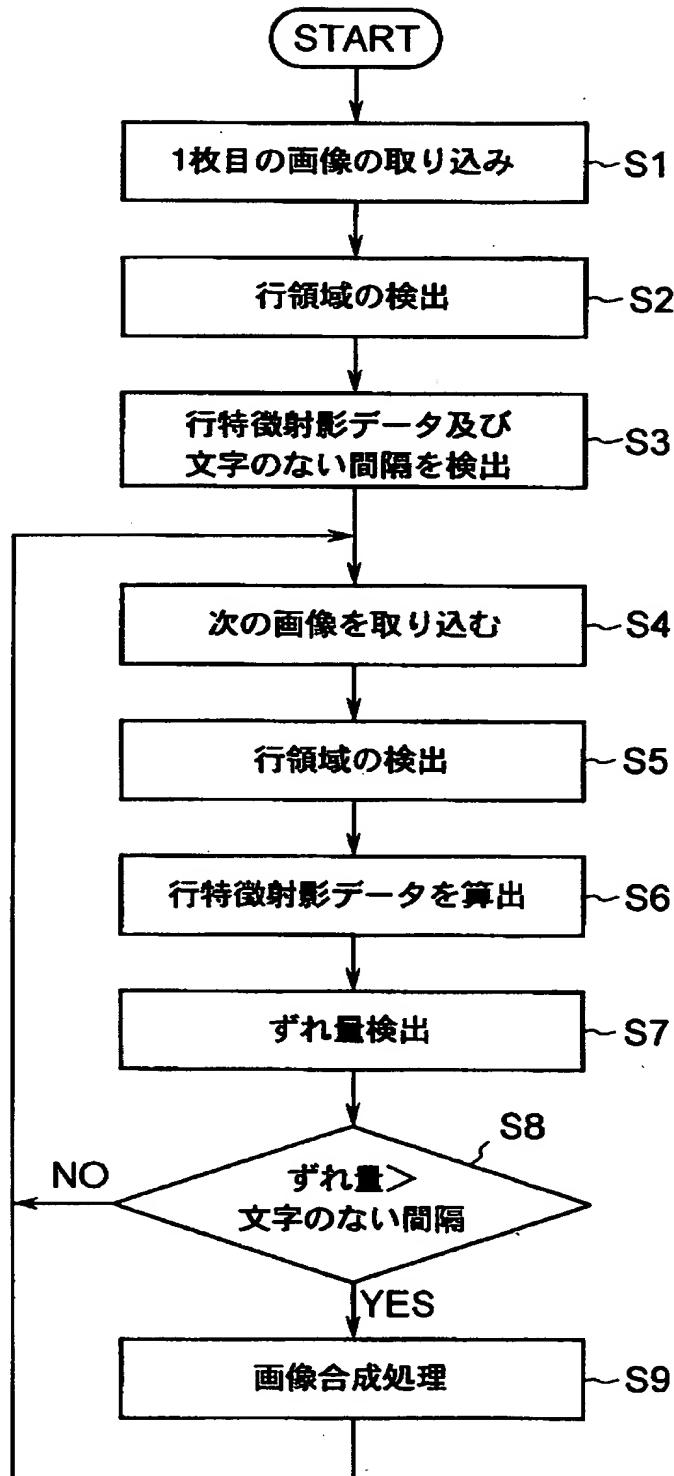
【図2】



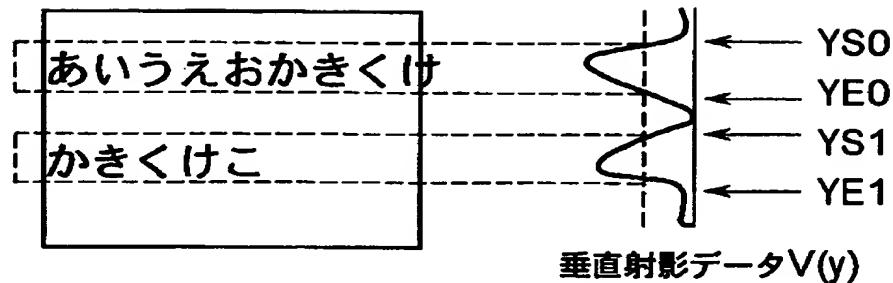
【図3】



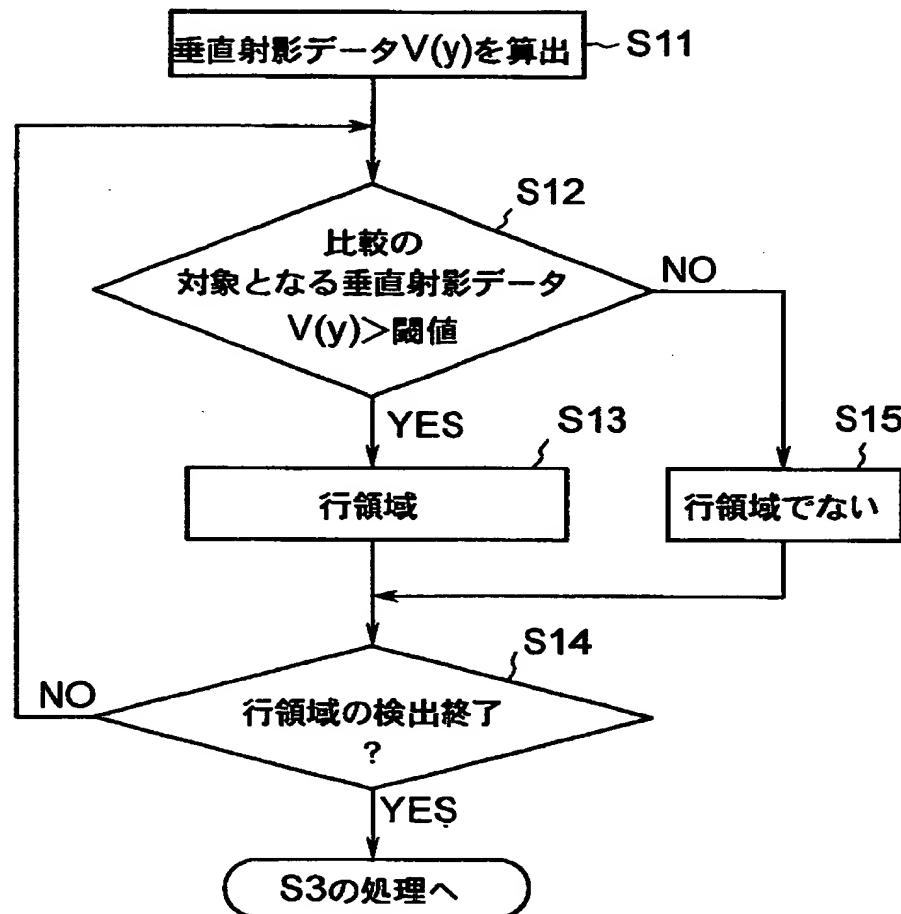
【図4】



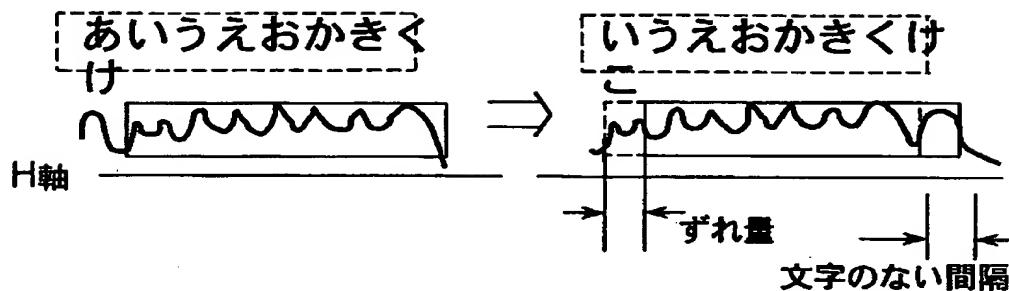
【図5】



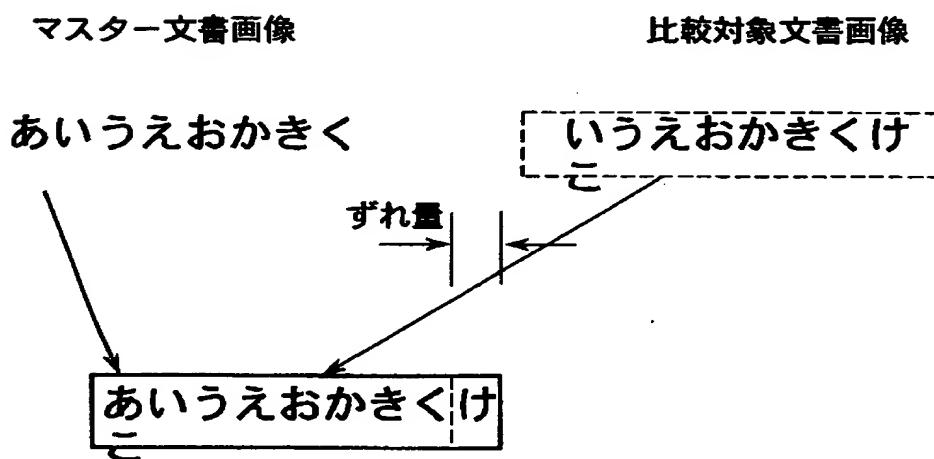
【図6】



【図7】

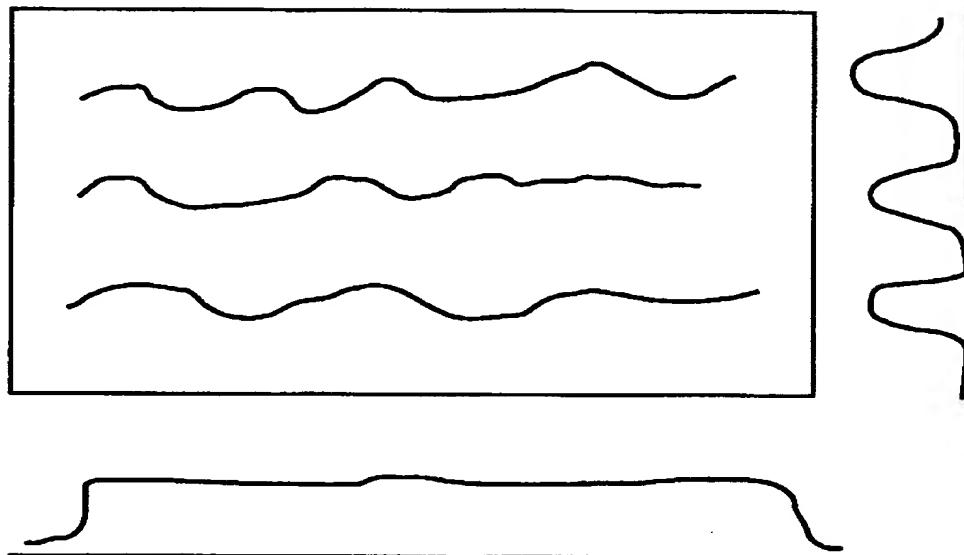


【図8】



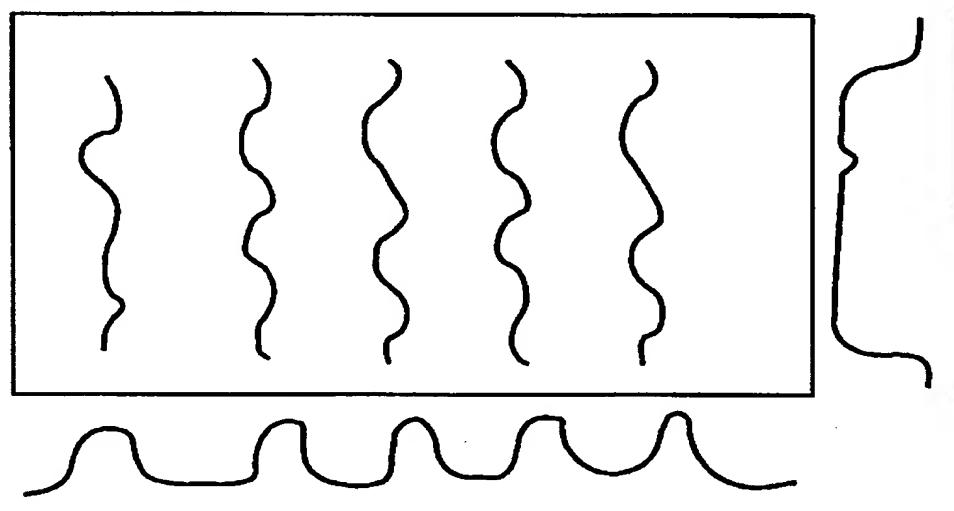
【図9】

横書きの場合

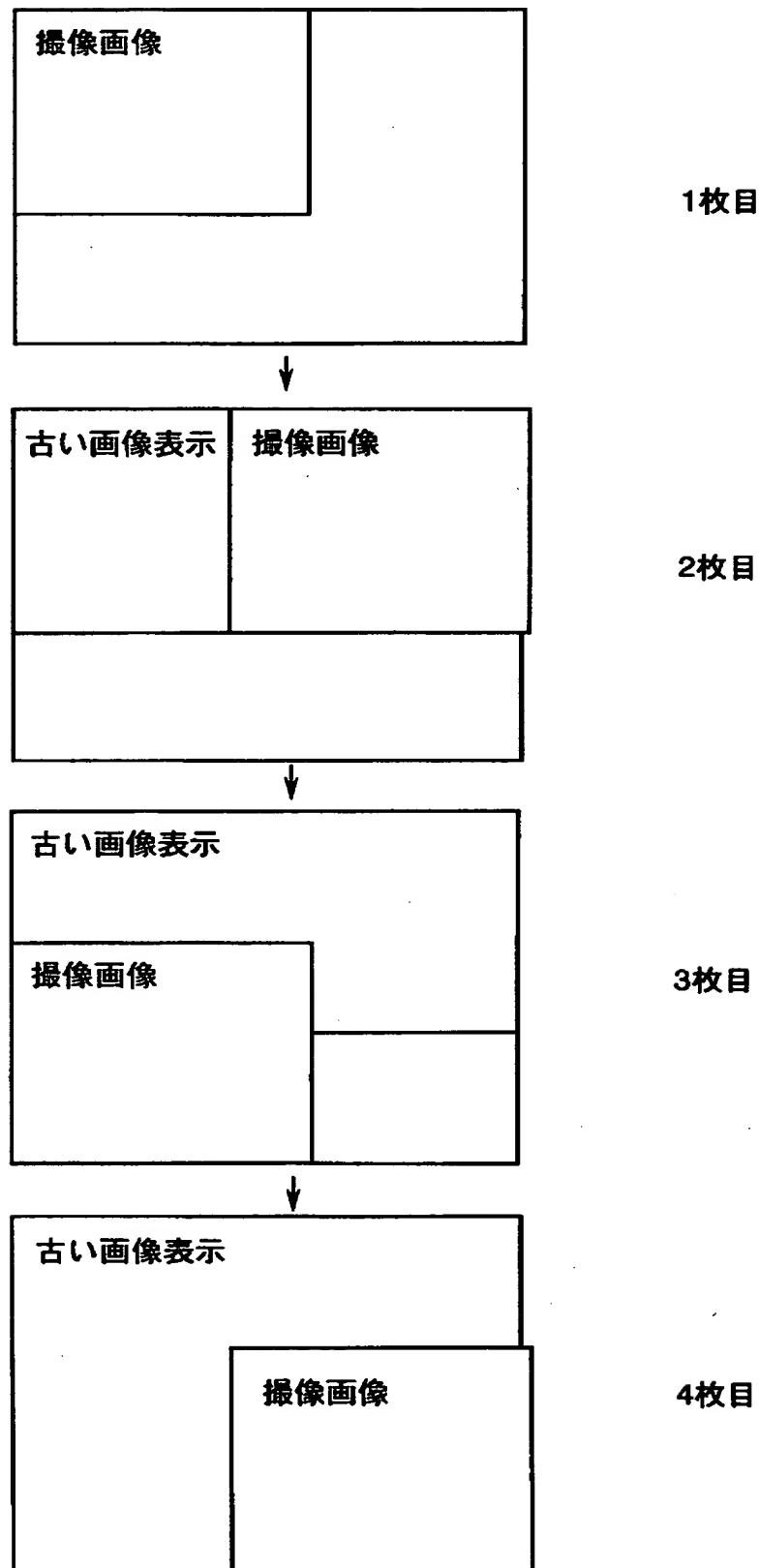


【図10】

縦書きの場合



【図11】



【図12】

これは例題の文章です
2行目のサンプルです
3つめはこれだ。

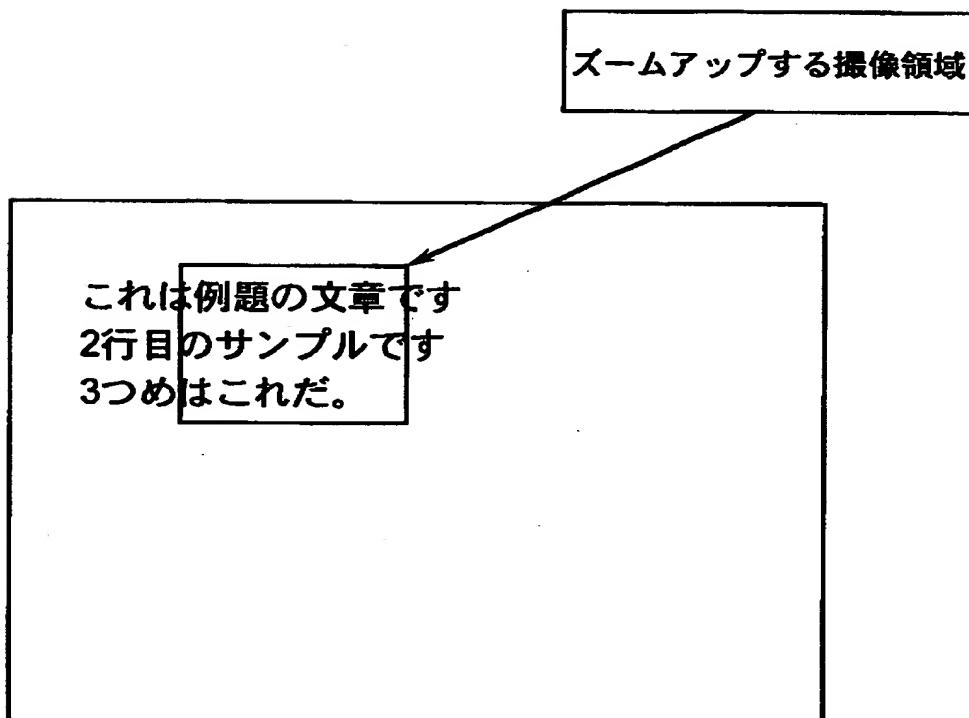
【図13】

これは例題の文章です
2行目のサンプルです
3つめはこれだ。

認識結果

誤認識

【図14】



【図15】

例題の文章
のサンプル
はこれだ。

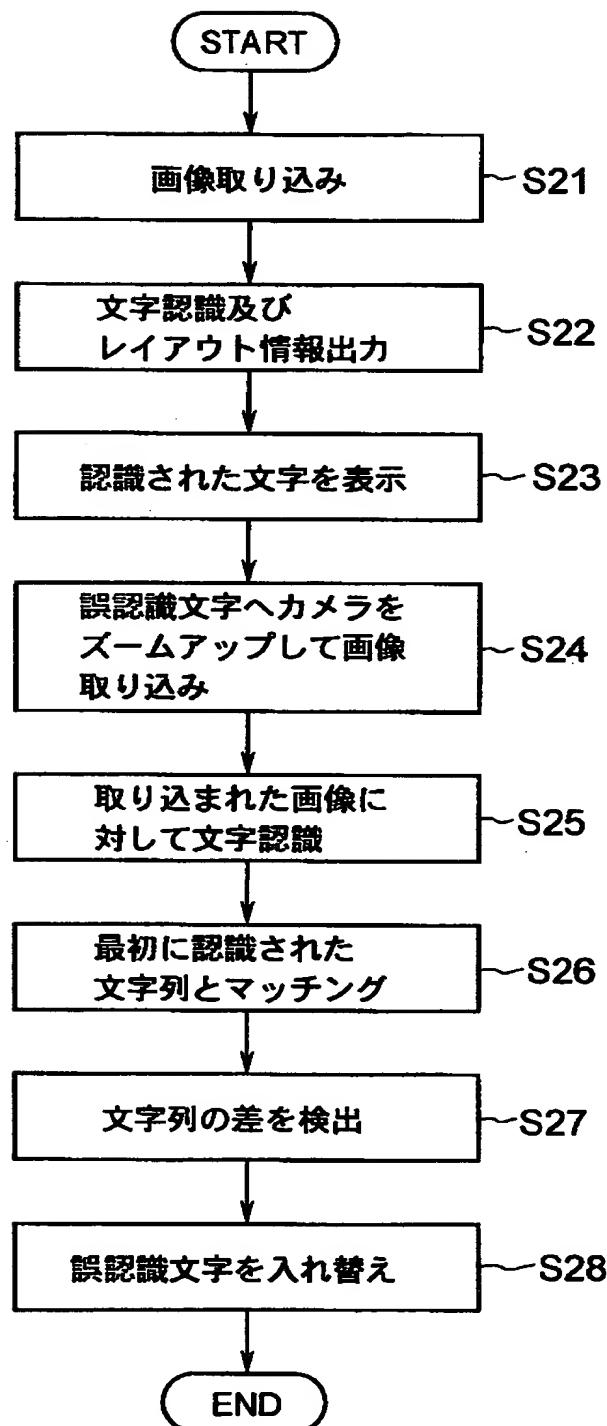
【図16】

例題の文章
のサンプル
はこれだ。 認識結果

↓
文字差し替えを行
う

これは例題の文章です
2行目のサンプルです
3つめはこれだ。

【図17】



【書類名】 要約書

【要約】

【課題】 カメラを動かすことで、広い部分の文書を高精度で読み取り、かつ誤認識部分の修正を容易に行なうことのできるカメラ画像認識装置を提供する。

【解決手段】 撮像された複数の被文書体の画像のうち特定の被文書体の画像の文字列の画像から、比較の対象となる被文書体の画像の文字列の画像のずれ量を撮像された被文書体の画像ごとに算出する。そして、算出されたずれ量が所定の量に達した場合に、特定の被文書体の画像の文字列の画像に、ずれ量が所定の値に達した被文書体の画像の文字列の画像のうち特定の被文書体の画像の文字列の画像に対して新たな文字の画像を張り合わせることにより文書画像を生成する。

【選択図】 図3

特2000-200241

出願人履歴情報

識別番号 [000003078]

1. 変更年月日 1990年 8月22日

[変更理由] 新規登録

住 所 神奈川県川崎市幸区堀川町72番地

氏 名 株式会社東芝